
Regras de Associação

Gabriel Fernandes Silva

Pre-processamento

Etapa anterior

Formato dos dados

- Não houve uma mudança significativa com relação ao pré-processamento feito na etapa anterior

#	Column	Non-Null Count		Dtype
0	Codigo da IES	1420	non-null	int64
1	Fim Lucrativo	1420	non-null	int64
2	Situacao	1420	non-null	int64
3	DIAS	1420	non-null	int64
4	Cursos Sem Ato 5 anos	1420	non-null	float64
5	Ultimo CI	1420	non-null	float64
6	Ultimo IGC	1420	non-null	float64
7	EAD_17	1420	non-null	int64
8	Variacao Matricula 16/17	1420	non-null	float64
9	Matriculas 17	1420	non-null	int64
10	% FIES	1420	non-null	float64
11	CURSOS	1420	non-null	int64
12	Saldo 2017	1420	non-null	float64
13	Variacao do Saldo 16/17	1420	non-null	float64

Transformação

Transformação dos dados para
aplicação da regra de associação

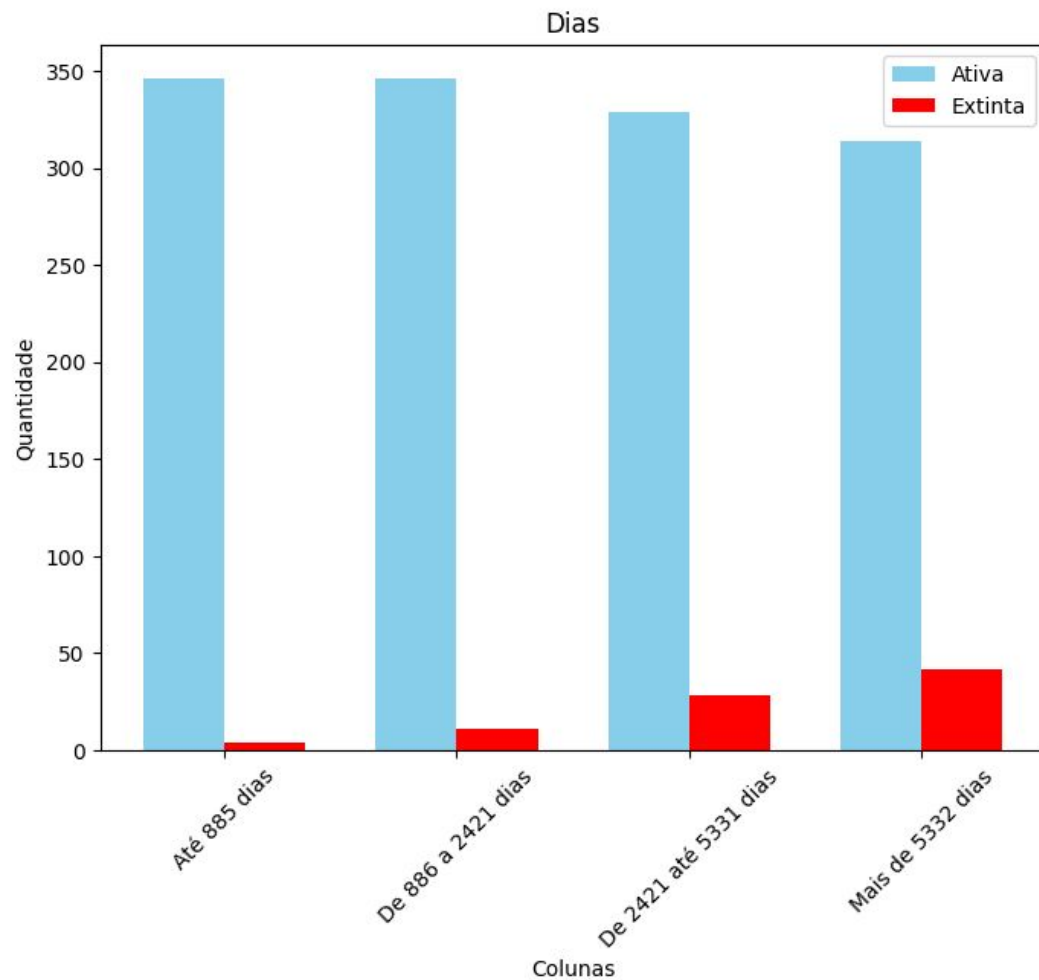


Dias

- A coluna de Dias foi dividida em faixas
 - Até 885 (até 25%)
 - De 886 a 2421 (25% a 50%)
 - De 2422 a 5331(50% a 75%)
 - Mais de 5331 (75%)

Até 885 dias	De 886 a 2421 dias	De 2421 até 5331 dias	Mais de 5332 dias
0	0	0	1
0	1	0	0
1	0	0	0
0	0	0	1
1	0	0	0

Dias

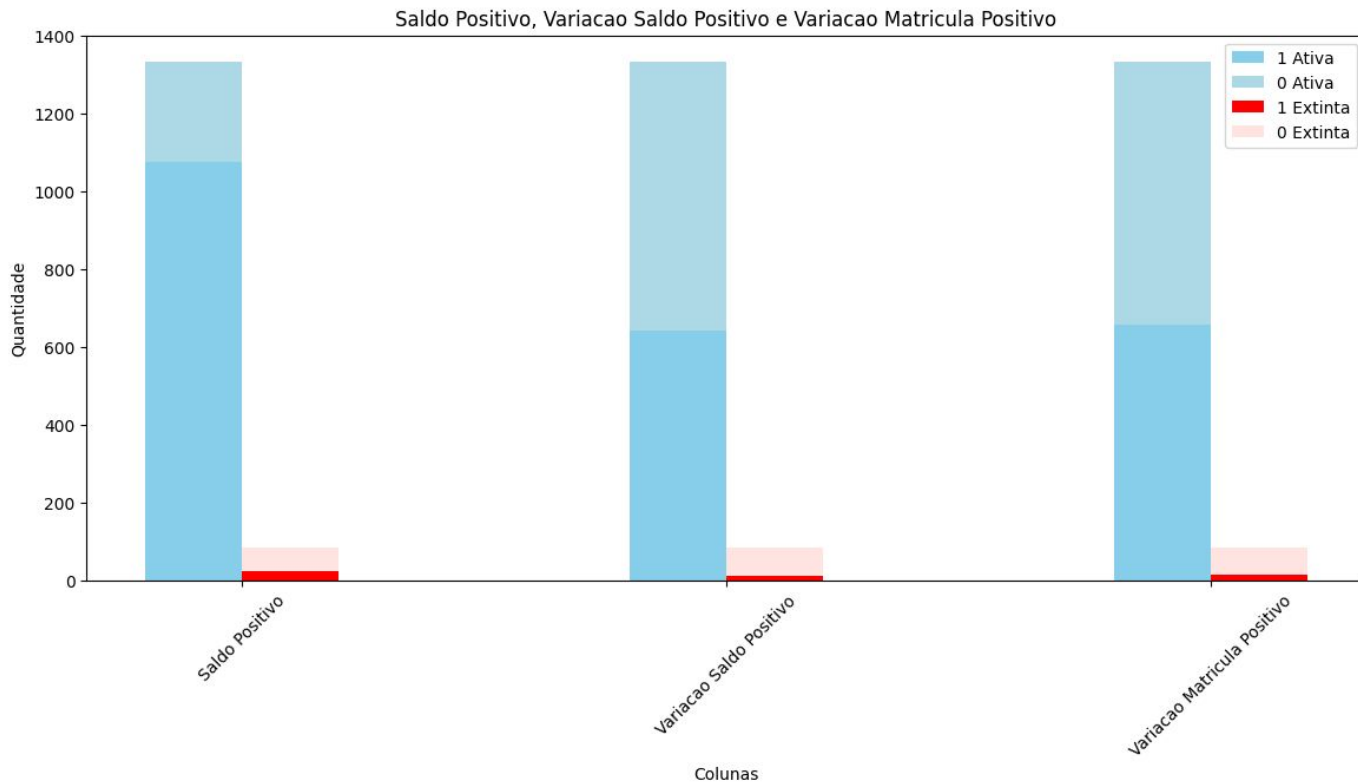


Saldo, Variação do Saldo e Variação de Matrícula

- A coluna de Saldo, Variação do Saldo e Variação da Matrícula como positivo ou não
 - Saldo Positivo
 - Variação Saldo Positivo
 - Variação Matrícula Positivo

Saldo Positivo	Variação Saldo Positivo	Variação Matrícula Positivo
1	0	1
1	1	0
1	0	0
1	0	1
0	1	1

Saldo, Variação do Saldo e Variação de Matrícula

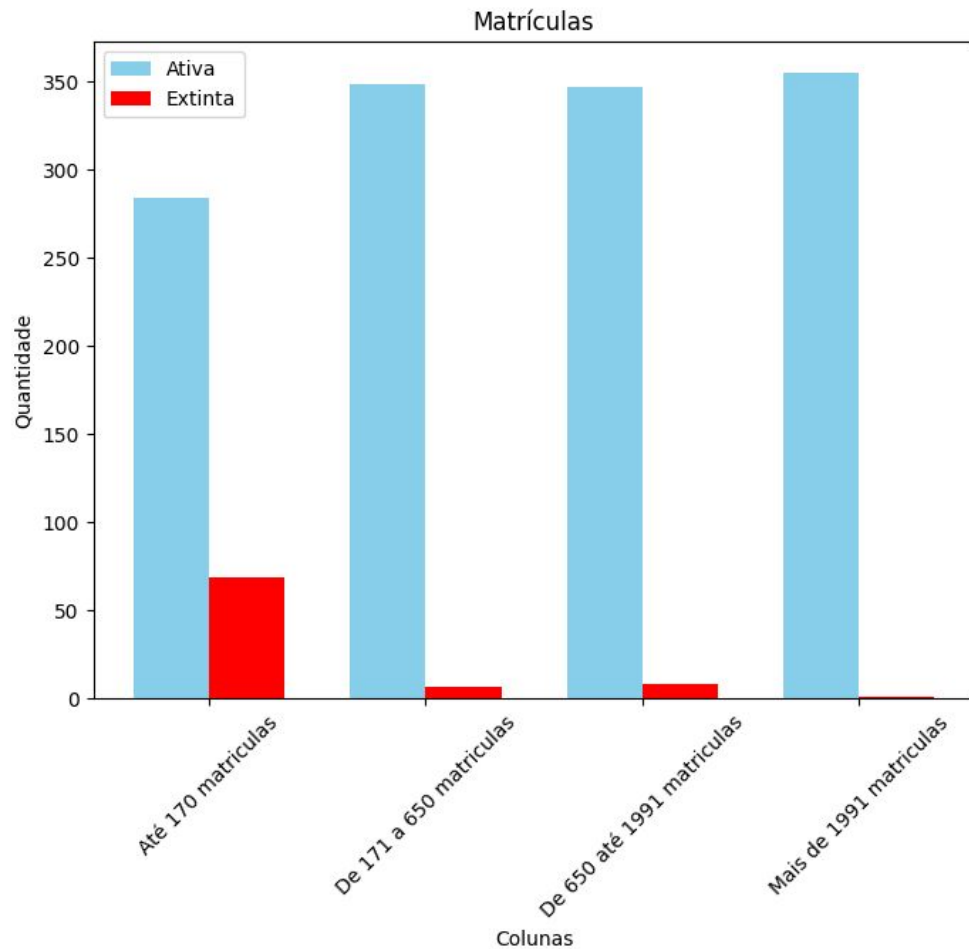


Matrículas

- A coluna de Matrículas foi dividida em faixas
 - Até 170 (até 25%)
 - De 171 a 650 (25% a 50%)
 - De 651 a 1991 (50% a 75%)
 - Mais de 1991 (75%)

Até 170 matriculas	De 171 a 650 matriculas	De 650 até 1991 matriculas	Mais de 1991 matriculas
0	1	0	0
0	1	0	0
0	0	0	1
0	0	1	0
1	0	0	0

Matrículas

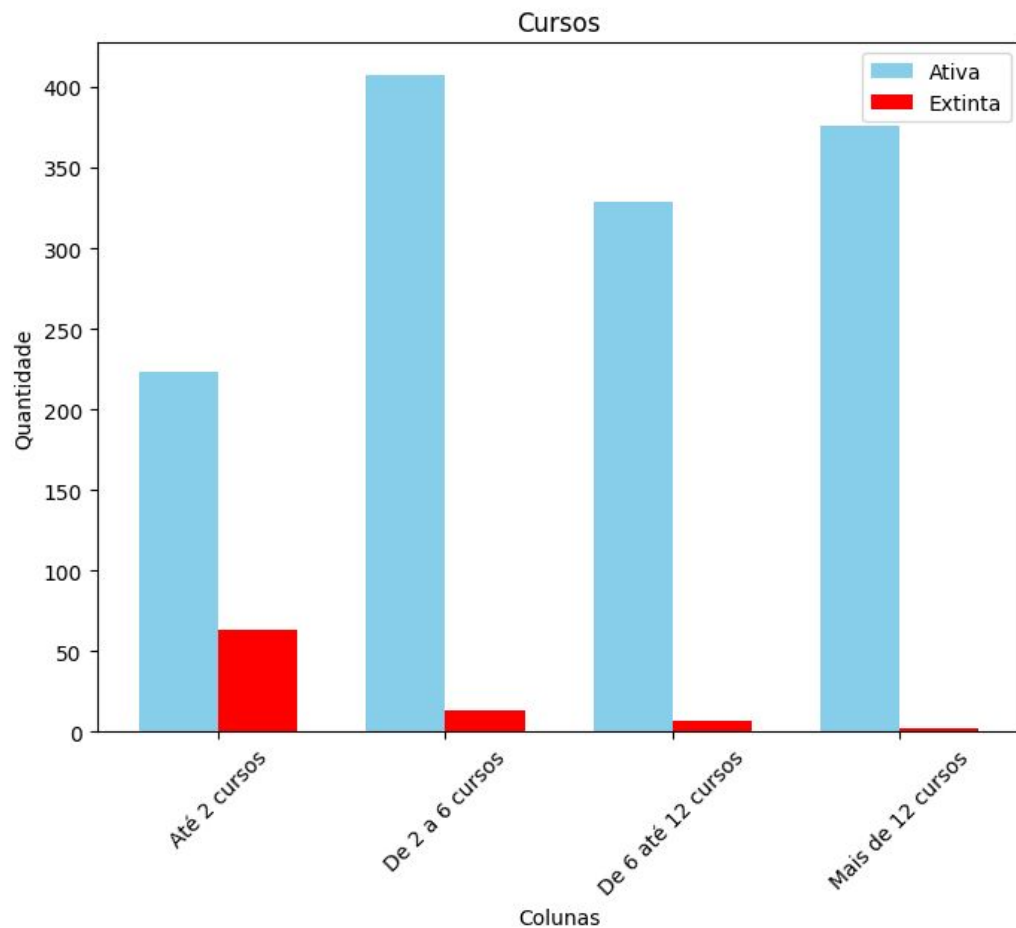


Cursos

- A coluna de Cursos foi dividida em faixas
 - Até 2 cursos (até 25%)
 - De 2 a 6 cursos (25% a 50%)
 - De 6 a 12 cursos (50% a 75%)
 - Mais de 12 cursos (75%)

Até 2 cursos	De 2 a 6 cursos	De 6 até 12 cursos	Mais de 12 cursos
0	1	0	0
0	1	0	0
0	0	0	1
0	0	1	0
1	0	0	0

Cursos

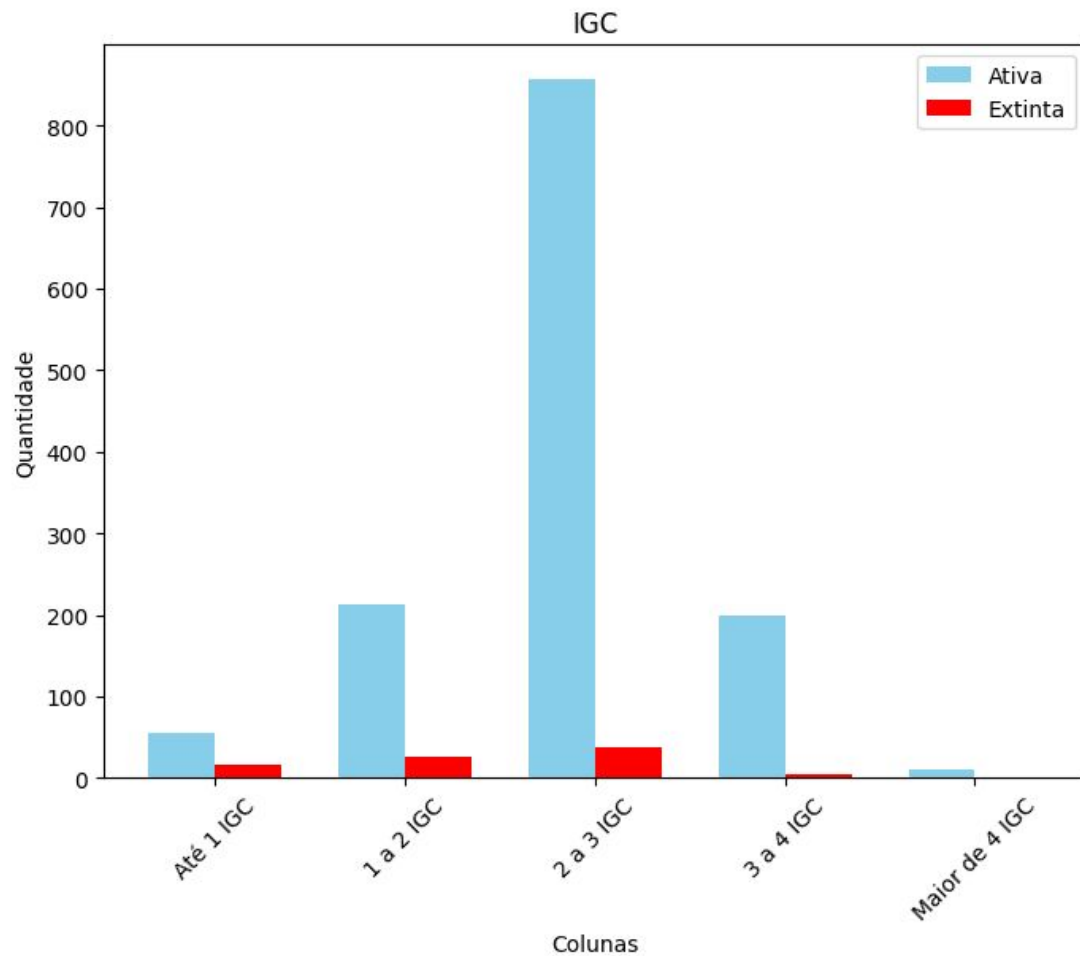


IGC (Último IGC)

- A coluna de Último IGC foi dividida em faixas
 - Até 1 IGC
 - De 1 a 2 IGC
 - De 2 a 3 IGC
 - De 3 a 4 IGC
 - Maior que 4 IGC

Até 1 IGC	1 a 2 IGC	2 a 3 IGC	3 a 4 IGC	Maior de 4
0	0	1	0	0
0	1	0	0	0
0	0	1	0	0
0	0	1	0	0
0	0	1	0	0

IGC

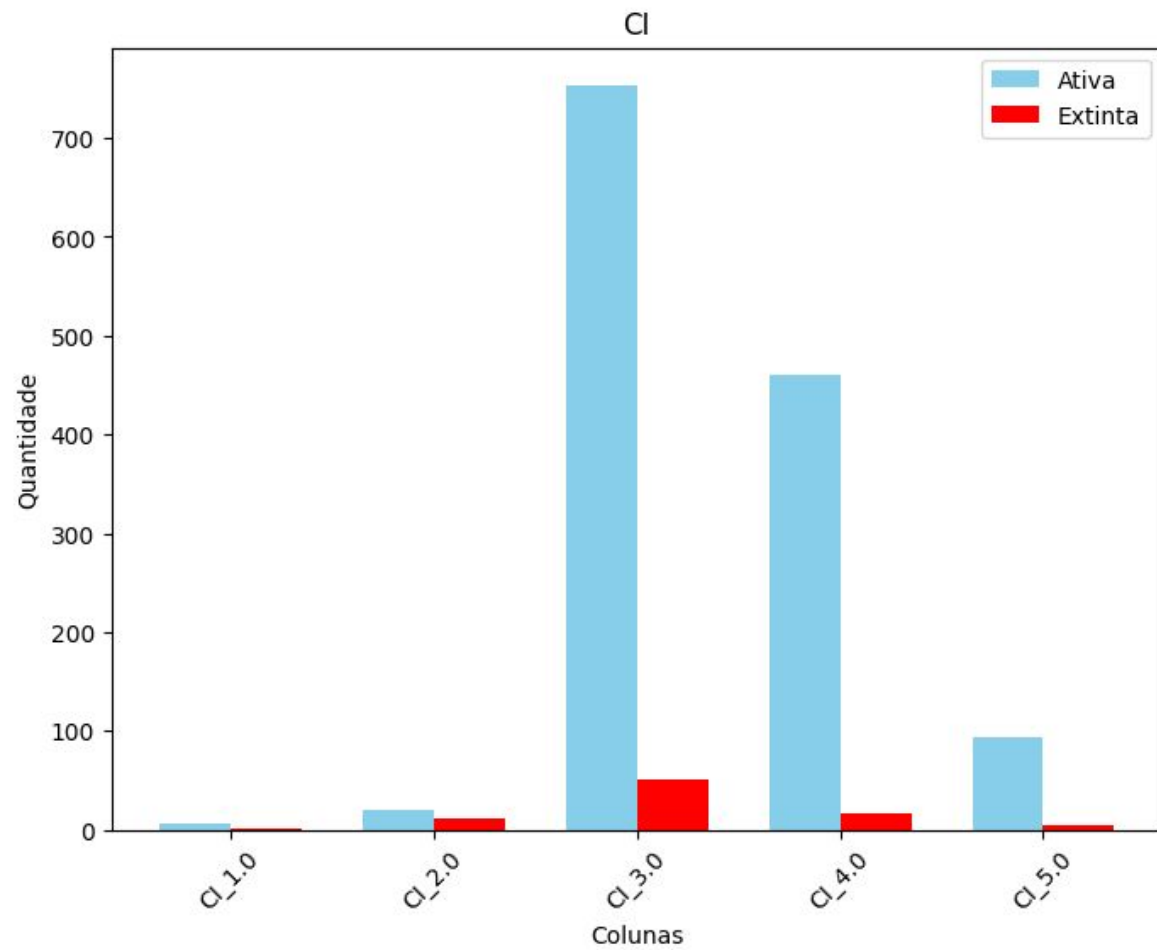


CI (Último CI)

- A coluna de Último CI foi dividida em faixas
 - CI 1
 - CI 2
 - CI 3
 - CI 4
 - CI 5

CI_1.0	CI_2.0	CI_3.0	CI_4.0	CI_5.0
0	0	1	0	0
0	0	1	0	0
0	0	0	1	0
0	0	1	0	0
0	0	1	0	0

CI

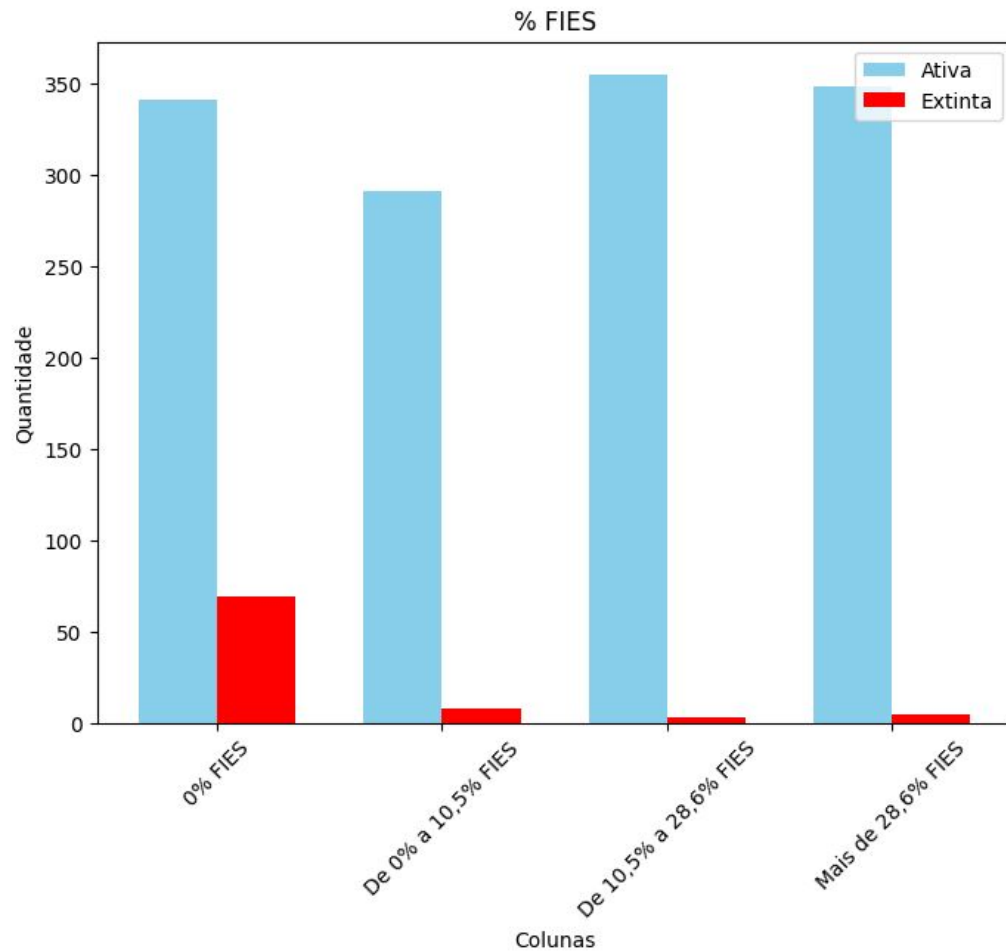


% FIES

- A coluna de % FIES foi dividida em faixas
 - 0 % FIES
 - De 0% a 10,5% FIES
 - De 10,5% a 28,6% FIES
 - Mais de 28,6% FIES

0% FIES	De 0% a 10,5% FIES	De 10,5% a 28,6% FIES	Mais de 28,6% FIES
1	0	0	0
0	1	0	0
0	1	0	0
0	0	1	0
1	0	0	0

% FIES

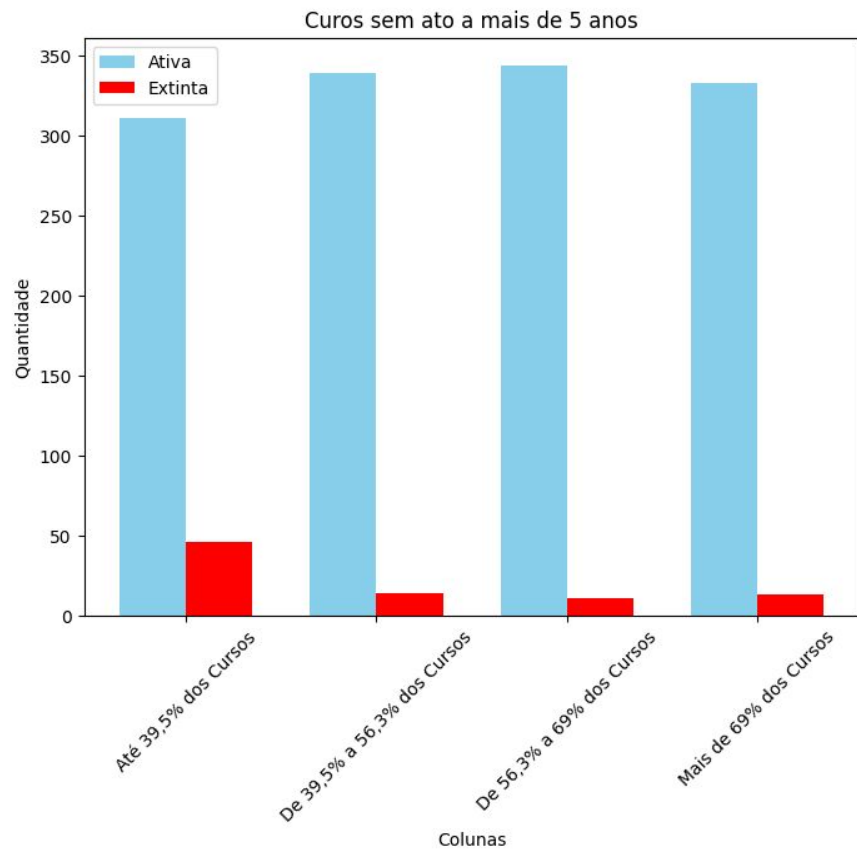


Cursos sem ato a mais de 5 anos

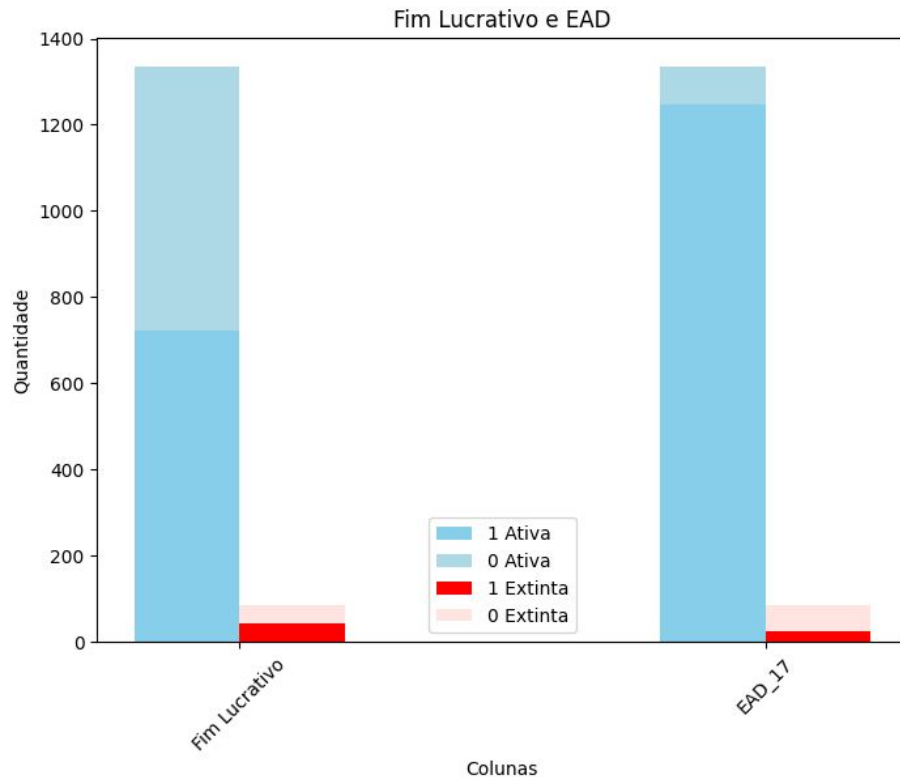
- A coluna de Cursos sem ato a mais de 5 anos foi dividida em faixas
 - Até 39,5% dos cursos (25%)
 - De 39,5% a 56,3% dos cursos (25% a 50%)
 - De 56,3% a 69% dos cursos (50% a 75%)
 - Mais de 69% dos cursos (75%)

Até 39,5% dos Cursos	De 39,5% a 56,3% dos Cursos	De 56,3% a 69% dos Cursos	Mais de 69% dos Cursos
0	0	1	0
0	0	0	1
0	0	1	0
0	0	1	0
0	0	0	1

Cursos sem ato a mais de 5 anos



Fim Lucrativo e EAD



Análise

Análise da regra de associação

APRIORI

Análise

Suporte Mínimo: 0.4

	support	itemsets
0	0.538028	(Fim Lucrativo)
1	0.940141	(Situacao)
2	0.895070	(EAD_17)
3	0.772535	(Saldo Positivo)
4	0.459155	(Variacao Saldo Positivo)
5	0.473944	(Variacao Matricula Positivo)
6	0.629577	(2 a 3 IGC)
7	0.566901	(CI_3.0)
8	0.508451	(Fim Lucrativo, Situacao)
9	0.486620	(Fim Lucrativo, EAD_17)
10	0.430986	(Fim Lucrativo, Saldo Positivo)
11	0.877465	(Situacao, EAD_17)
12	0.757042	(Situacao, Saldo Positivo)
13	0.451408	(Situacao, Variacao Saldo Positivo)
14	0.463380	(Situacao, Variacao Matricula Positivo)
15	0.602817	(Situacao, 2 a 3 IGC)
16	0.530986	(Situacao, CI_3.0)
17	0.772535	(Saldo Positivo, EAD_17)
18	0.459155	(Variacao Saldo Positivo, EAD_17)
19	0.473944	(Variacao Matricula Positivo, EAD_17)
20	0.594366	(2 a 3 IGC, EAD_17)
21	0.493662	(CI_3.0, EAD_17)
22	0.412676	(Variacao Saldo Positivo, Saldo Positivo)
23	0.516197	(2 a 3 IGC, Saldo Positivo)
24	0.417606	(Saldo Positivo, CI_3.0)

25	0.476761	(Fim Lucrativo, Situacao, EAD_17)
26	0.422535	(Fim Lucrativo, Saldo Positivo, Situacao)
27	0.430986	(Fim Lucrativo, Saldo Positivo, EAD_17)
28	0.757042	(Situacao, Saldo Positivo, EAD_17)
29	0.451408	(Situacao, Variacao Saldo Positivo, EAD_17)
30	0.463380	(Situacao, Variacao Matricula Positivo, EAD_17)
31	0.584507	(Situacao, EAD_17, 2 a 3 IGC)
32	0.483099	(Situacao, CI_3.0, EAD_17)
33	0.406338	(Situacao, Variacao Saldo Positivo, Saldo Posi...
34	0.508451	(Situacao, Saldo Positivo, 2 a 3 IGC)
35	0.409155	(Situacao, Saldo Positivo, CI_3.0)
36	0.412676	(Variacao Saldo Positivo, Saldo Positivo, EAD_17)
37	0.516197	(2 a 3 IGC, Saldo Positivo, EAD_17)
38	0.417606	(Saldo Positivo, CI_3.0, EAD_17)
39	0.422535	(Fim Lucrativo, Saldo Positivo, Situacao, EAD_17)
40	0.406338	(Situacao, Variacao Saldo Positivo, Saldo Posi...
41	0.508451	(Situacao, Saldo Positivo, EAD_17, 2 a 3 IGC)
42	0.409155	(Situacao, Saldo Positivo, CI_3.0, EAD_17)

Associação

index	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
46	frozenset({'Variacao Saldo Positivo', 'Saldo Positivo'})	frozenset({'Situacao'})	0.4126760563380282	0.9401408450704225	0.4063380281690141	0.984641638225256	1.0473341769886617
68	frozenset({'Variacao Saldo Positivo', 'Saldo Positivo', 'EAD_17'})	frozenset({'Situacao'})	0.4126760563380282	0.9401408450704225	0.4063380281690141	0.984641638225256	1.0473341769886617
40	frozenset({'2 a 3 IGC', 'EAD_17'})	frozenset({'Situacao'})	0.5943661971830986	0.9401408450704225	0.5845070422535211	0.9834123222748815	1.0460265899852672
34	frozenset({'Variacao Saldo Positivo', 'EAD_17'})	frozenset({'Situacao'})	0.4591549295774648	0.9401408450704225	0.45140845070422536	0.9831288343558282	1.045725052273615
7	frozenset({'Variacao Saldo Positivo'})	frozenset({'Situacao'})	0.4591549295774648	0.9401408450704225	0.45140845070422536	0.9831288343558282	1.045725052273615
22	frozenset({'Fim Lucrativo', 'Saldo Positivo'})	frozenset({'Situacao'})	0.4309859154929577	0.9401408450704225	0.4225352112676056	0.9803921568627452	1.0428141293970772
61	frozenset({'Fim Lucrativo', 'Saldo Positivo', 'EAD_17'})	frozenset({'Situacao'})	0.4309859154929577	0.9401408450704225	0.4225352112676056	0.9803921568627452	1.0428141293970772
6	frozenset({'Saldo Positivo'})	frozenset({'Situacao'})	0.7725352112676056	0.9401408450704225	0.7570422535211268	0.9799453053783045	1.0423388266945262
51	frozenset({'Saldo Positivo', 'CI_3.0'})	frozenset({'Situacao'})	0.4176056338028169	0.9401408450704225	0.4091549295774648	0.9797639123102867	1.0421458842551365
82	frozenset({'EAD_17', 'Saldo Positivo', 'CI_3.0'})	frozenset({'Situacao'})	0.4176056338028169	0.9401408450704225	0.4091549295774648	0.9797639123102867	1.0421458842551365
20	frozenset({'Fim Lucrativo', 'EAD_17'})	frozenset({'Situacao'})	0.4866197183098592	0.9401408450704225	0.47676056338028167	0.979739507959479	1.0421199260692586
43	frozenset({'CI_3.0', 'EAD_17'})	frozenset({'Situacao'})	0.4936619718309859	0.9401408450704225	0.4830985915492958	0.978601997146933	1.0409099894746403
37	frozenset({'Variacao Matricula Positivo', 'EAD_17'})	frozenset({'Situacao'})	0.47394366197183097	0.9401408450704225	0.46338028169014084	0.9777117384843983	1.0399630476762889
8	frozenset({'Variacao Matricula Positivo'})	frozenset({'Situacao'})	0.47394366197183097	0.9401408450704225	0.46338028169014084	0.9777117384843983	1.0399630476762889
74	frozenset({'2 a 3 IGC', 'Saldo Positivo', 'EAD_17'})	frozenset({'Situacao'})	0.5161971830985915	0.9401408450704225	0.5084507042253521	0.984993178717599	1.0477081002089816
48	frozenset({'2 a 3 IGC', 'Saldo Positivo'})	frozenset({'Situacao'})	0.5161971830985915	0.9401408450704225	0.5084507042253521	0.984993178717599	1.0477081002089816
9	frozenset({'2 a 3 IGC'})	frozenset({'Situacao'})	0.6295774647887324	0.9401408450704225	0.6028169014084507	0.9574944071588366	1.018458470535991
0	frozenset({'Fim Lucrativo'})	frozenset({'Situacao'})	0.5380281690140845	0.9401408450704225	0.5084507042253521	0.9450261780104713	1.0051963841010256
10	frozenset({'CI_3.0'})	frozenset({'Situacao'})	0.5669014084507042	0.9401408450704225	0.5309859154929577	0.9366459627329191	0.9962825970642286

Análise

Suporte Mínimo: 0.5

	support	itemsets
0	0.538028	(Fim Lucrativo)
1	0.940141	(Situacao)
2	0.895070	(EAD_17)
3	0.772535	(Saldo Positivo)
4	0.629577	(2 a 3 IGC)
5	0.566901	(CI_3.0)
6	0.508451	(Fim Lucrativo, Situacao)
7	0.877465	(EAD_17, Situacao)
8	0.757042	(Saldo Positivo, Situacao)
9	0.602817	(2 a 3 IGC, Situacao)
10	0.530986	(CI_3.0, Situacao)
11	0.772535	(EAD_17, Saldo Positivo)
12	0.594366	(EAD_17, 2 a 3 IGC)
13	0.516197	(Saldo Positivo, 2 a 3 IGC)
14	0.757042	(EAD_17, Saldo Positivo, Situacao)
15	0.584507	(EAD_17, 2 a 3 IGC, Situacao)
16	0.508451	(2 a 3 IGC, Saldo Positivo, Situacao)
17	0.516197	(EAD_17, Saldo Positivo, 2 a 3 IGC)
18	0.508451	(EAD_17, 2 a 3 IGC, Saldo Positivo, Situacao)

Associação

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
20	(Saldo Positivo, 2 a 3 IGC)	(Situacao)	0.516197	0.940141	0.508451	0.984993	1.047708	0.023153	3.988796	0.094120
26	(EAD_17, Saldo Positivo, 2 a 3 IGC)	(Situacao)	0.516197	0.940141	0.508451	0.984993	1.047708	0.023153	3.988796	0.094120
17	(EAD_17, 2 a 3 IGC)	(Situacao)	0.594366	0.940141	0.584507	0.983412	1.046027	0.025719	3.608652	0.108476
13	(Saldo Positivo, EAD_17)	(Situacao)	0.772535	0.940141	0.757042	0.979945	1.042339	0.030750	2.984795	0.178573
4	(Saldo Positivo)	(Situacao)	0.772535	0.940141	0.757042	0.979945	1.042339	0.030750	2.984795	0.178573
5	(2 a 3 IGC)	(Situacao)	0.629577	0.940141	0.602817	0.957494	1.018458	0.010925	1.408265	0.048928
0	(Fim Lucrativo)	(Situacao)	0.538028	0.940141	0.508451	0.945026	1.005196	0.002628	1.088867	0.011190
6	(Cl_3.0)	(Situacao)	0.566901	0.940141	0.530986	0.936646	0.996283	-0.001981	0.944836	-0.008542
2	(EAD_17)	(Situacao)	0.895070	0.940141	0.877465	0.980330	1.042748	0.035973	3.043239	0.390700

Análise

Suporte Mínimo: 0.6

	support	itemsets
0	0.940141	(Situacao)
1	0.895070	(EAD_17)
2	0.772535	(Saldo Positivo)
3	0.629577	(2 a 3 IGC)
4	0.877465	(Situacao, EAD_17)
5	0.757042	(Saldo Positivo, Situacao)
6	0.602817	(2 a 3 IGC, Situacao)
7	0.772535	(Saldo Positivo, EAD_17)
8	0.757042	(Saldo Positivo, Situacao, EAD_17)

Associação

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
1	(EAD_17)	(Situacao)	0.895070	0.940141	0.877465	0.980330	1.042748	0.035973	3.043239	0.390700
2	(Saldo Positivo)	(Situacao)	0.772535	0.940141	0.757042	0.979945	1.042339	0.030750	2.984795	0.178573
8	(Saldo Positivo, EAD_17)	(Situacao)	0.772535	0.940141	0.757042	0.979945	1.042339	0.030750	2.984795	0.178573
4	(2 a 3 IGC)	(Situacao)	0.629577	0.940141	0.602817	0.957494	1.018458	0.010925	1.408265	0.048928

Resultado

Resultados obtidos tomando
situação como consequente

Resultados

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
20	(Saldo Positivo, 2 a 3 IGC)	(Situacao)	0.516197	0.940141	0.508451	0.984993	1.047708	0.023153	3.988796	0.094120
26	(EAD_17, Saldo Positivo, 2 a 3 IGC)	(Situacao)	0.516197	0.940141	0.508451	0.984993	1.047708	0.023153	3.988796	0.094120
17	(EAD_17, 2 a 3 IGC)	(Situacao)	0.594366	0.940141	0.584507	0.983412	1.046027	0.025719	3.608652	0.108476
2	(EAD_17)	(Situacao)	0.895070	0.940141	0.877465	0.980330	1.042748	0.035973	3.043239	0.390700

Principais antecessores:

- Saldo Positivo
- 2 a 3 IGC
- EAD

Obrigado!
